DOCUMENT RESUME

ED 164 563 TM 007 678

AUTHOR Folsom, R. E.; Weber, J. H.
TITLE Relative Efficiencies of a Three-Stage Versus a Two-Stage Sample Design For a New NLS Cohort Study. 22U-884-38.
INSTITUTION Research Triangle Inst., Research Triangle Park, N.C. Center for Sampling Research and Design.
PUB DATE Mar 76
NOTE 67p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Attrition (Research Studies); Cost Effectiveness; *Costs; Efficiency; *Graduate Surveys; *Longitudinal Studies; National Surveys; *Research Design; Research Methodology; *Sampling; Senior High Schools; Statistical Analysis
IDENTIFIERS *National Longitudinal Study

ABSTRACT

Two sampling designs were compared for the planned 1978 national longitudinal survey of high school seniors with respect to statistical efficiency and cost. The 1972 survey used a stratified two-stage sample of high schools and seniors within schools. In order to minimize interviewer travel costs, an alternate sampling design was proposed, clustering schools in primary sampling units (PSUs). Two models were compared for efficiency with either 4 or 5 schools per PSU, and 9 seniors per school, and about 2,500 schools. The statistical procedures and assumptions needed for the efficiency estimates are given. The cost model includes total valuable costs for the base-year survey and five follow-up surveys. Cost depends on the number of students to be selected and on the number of PSUs. It was assumed that 40 percent of the students would move out of the PSU before the first follow-up survey. Costs were also estimated for either 30 percent or 50 percent moving out. It was concluded that gains in precision are not sufficient to justify redesigning the sample. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164563

22U-884-38

RELATIVE EFFICIENCIES OF A THREE-STAGE VERSUS A TWO-STAGE SAMPLE DESIGN FOR A NEW NLS COHORT STUDY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

by

R. E. Folsom
J. H. Weber

Sampling Research and Design Center
Research Triangle Institute
Research Triangle Park, N.C. 27709

March 1976

TM007 678

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

TABLE OF CONTENTS

		Page
LIST OF TABLES		iii
I.	INTRODUCTION	1
II.	THE TWO- AND THREE-STAGE VARIANCE COMPONENT MODEL	6
III.	THE EFFICIENCY CRITERION TO COMPARE A THREE-STAGE VERSUS A TWO-STAGE SAMPLE DESIGN	11
IV.	COST MODELS	20
	A. Distance Traveled Model	20
	B. Cost of a Single Follow-up Survey Using a Three-stage Sample Design	24
	C. Costs of All Surveys in an NLS Cohort	27
	D. An Alternative Procedure in Determining Both the Two- and Three-Stage Cost Models	30
	E. Cost Models for Two Proportions of Students Moving and Two Methods of Allocating Costs	32
V.	RESULTS OF COMPARING THE RELATIVE EFFICIENCIES OF A THREE- VERSUS A TWO-STAGE SAMPLE DESIGN	42
VI.	CONCLUSIONS	46
REFERENCES		48
APPENDIX A - SUMMARY STATEMENTS OF 1972 NLS AND 1974 NAEP DESIGN		49
APPENDIX B - TABLES FOR DETERMINING COST COMPONENTS		55

LIST OF TABLES

		Page
1.	Efficiencies of the Optimum Three-stage Design Relative to the Optimum Two-stage Design	5
2.	Average Values of the NLS Change-year Variance Component Ratios	15
3.	Average Values of the Three-stage NAEP Variance Component Ratios	18
4.	Estimated Direct Charges Allocated to PSU, School, and Student Costs	28
5.	Calculation of Estimated Variable Cost Parameters, $\$_1$, $\$_2$, and $\$_3$ for an NLS Follow-up Survey	29
6.	Optimum Number of Schools and Students per School Selected for Each Domain Variable Subject to the Cost Model, $C-C_0 = \$361\ n_1 + \$180\ n_1n_2$	36
7.	Optimum Number of PSUs, Schools per PSU, and Students per School Selected for Each Domain Variable Subject to the Cost Model, $\$-\$_0 = \$325\ m_1 + \$130\ m_1m_2 + \$174\ m_1m_2m_3$	37
8.	Optimum Number of Schools and Students per School Selected for Each Domain Variable Subject to the Cost Model, $C-C_0 = \$567\ n_1 + \$169\ n_1n_2$	38
9.	Optimum Number of PSUs, Schools per PSU, and Students per School Selected for Each Domain Variable Subject to the Cost Model, $\$-\$_0 = \$599\ m_1 + \$130\ m_1m_2 + \$157\ m_1m_2m_3$	39
10.	Relative Efficiencies of the "Best" Three-stage Design Relative to the "Best" Two-stage Design Assuming the Optimum Design Parameters for the "All Statistics" Group and Cost Procedure A	40
11.	Relative Efficiencies of the "Best" Three-stage Design Relative to the "Best" Two-Stage Design Assuming the Optimum Design Parameters for the "All Statistics" Group and Cost Procedure B.	41

LIST OF TABLES (Continued)

		Page
B-1.	Estimated Direct Charges Allocated to School and Student Costs, NLS Base-year Survey, Two-Stage Design	56
B-2.	Estimated Man-months of Direct Labor Allocated to School and Student Costs, NLS Base Year, Two-stage Design	56
B-3.	Calculation of Estimated Variable Cost Parameters C_1 and C_2, NLS Base-year Survey, Two-stage Design	57
B-4.	Estimated Direct Charges Allocated to PSU, School, and Student Costs	58
B-5.	Calculation of Estimated Variable Cost Parameters, C_1 and C_2 for an NLS Follow-up Survey using a Two-Stage Sampling Design .	59
B-6.	Calculation of Estimated Variable Cost Parameters, $\$_1$, $\$_2$, and $\$_3$ for an NLS Follow-up Survey using a Three-stage Sampling Design	60
B-7.	Estimated Direct Charges Allocated to School and Student Costs, NLS Follow-up Survey, Two-stage Design	61
B-8.	Estimated Man-months of Direct Labor, by Task, by Position, NLS Follow-up Survey, Two-stage design	61
B-9.	Calculation of Estimated Variable Cost Parameters, C_1 and C_2, for an NLS Follow-up Survey, Two-stage Design	62

I. INTRODUCTION

The National Center for Educational Statistics began planning in 1969 for a series of national longitudinal surveys of high school graduates. The basic purpose of these surveys was to relate high school educational experiences and other student characteristics to subsequent educational and vocational activities. The first cohort of 1972 high school seniors was selected using a deeply stratified two-stage sample of high schools and seniors within schools. Base-year (BY) data for this first cohort was collected in school during the spring of 1972; first follow-up (FFU) data was obtained by a combination of mail and personal interviews from October 1973 through April of 1974. Second follow-up responses were collected in a similar fashion during the corresponding period of 1974 and 1975. Third follow-up data collection is scheduled to begin in October of 1976, and a fourth follow-up has been proposed for 1979-1980. Meanwhile, plans are underway to induct the second cohort of high school seniors in the spring of 1978. In a previous NLS sample design efficiency report, Moore and Shah [3] have estimated the optimum numbers of schools and seniors per school for a new cohort sample allowing for up to five follow-up surveys. As one considers the interviewer travel costs associated with multiple follow-up surveys, it becomes apparent that a three-stage design clustering schools in primary sampling units (PSUs) of one or more counties might in the long run be a cost effective alternative. This report compares the efficiency of a deeply stratified three-stage design patterned after RTI's National Assessment of Educational Progress (NAEP) sample to an optimized

version of the 1972 two-stage design. To illustrate more precisely the nature of the two- and three-stage designs being compared, appendix A provides statements of the NLS 1972 cohort and the NAEP Year 05 (1974) sample designs.

The precision of an optimized NLS two-stage design was modeled using the change year variance components introduced in the Moore and Shah report alluded to previously. For the competing three-stage design, NLS two-stage variance components were partitioned into components for PSUs, and schools within PSUs. This partitioning was accomplished using ratios of NAEP Year 05 17-year-old variance components. The two- and three-stage variance models are presented in chapter II along with a description of the partitioning scheme used to approximate three-stage variance components.

An efficiency criterion for comparing the precision of cost equivalent two- and three-stage sample designs across a range of NLS response variables and reporting subgroups is developed in chapter III. Crucial ingredients in the resulting measures of relative efficiency are linear cost components for the competing two- and three-stage designs. The two-stage cost components used here were those developed by Moore and Shah for a base-year survey and five follow-up surveys. An interviewer travel model was then structured for the three-stage sample. Translating distance traveled into interviewer salaries and expenses, three-stage cost components were then estimated. These cost considerations are summarized in chapter IV.

Recognizing the subjective elements in the cost modeling process and noting the limitations in available data appropriate to this study, alternate cost models for both the two- and three-stage designs were explored. Two different procedures were used to allocate survey specialist salaries and

travel costs to the various components in the model. The form of interviewer travel model used to estimate cost components depends on the percentage of sample students who would move from the PSU in which their school was located to another area outside their initial PSU. A mobility study is underway to determine the number of students in the first NLS cohort who moved from the county in which their school was located between the base year and first follow-up and, between the first and second follow-up. The results of this study will be presented in a subsequent report. Lacking more exact figures on the fraction of movers, three mobility rates were assumed; namely 30 percent, 40 percent, and 50 percent. Six alternative models were then obtained by using the two methods of allocating costs in combination with the three values for the proportion of students moving.

The optimum design parameters (number of PSUs, schools per PSU, and students per school for the three-stage design and numbers of schools and students per school for the two-stage design) were calculated for all of the domains considered subject to the six different cost constraints. These design parameters are optimal in the sense of minimum variance for a particular domain at a fixed price. The average relative efficiency of the three-stage versus the two-stage design using the optimum "all statistics" design parameters was determined for each of 6 cost models and for each of 12 domain and type of statistic groupings. These results are detailed in chapter V. The major findings are indicated below.

The proportion of students moving had very little effect on the relative efficiencies. The two methods of allocating survey specialist costs had a much greater impact. The three-stage design for which all survey specialist

costs were allocated to the student level had a smaller variance than the cost equivalent two-stage sample for seven of the eight specific domains examined with the exception being "father's education less than high school." The reduction in variance for this cost allocation scheme was 5 percent for national means and from zero to 5 percent for specific domain means; over "all statistics," the three-stage relative efficiency gains averaged one percent. The second model which allocated a portion of the survey specialist costs to PSUs and schools made the three-stage design look a bit more favorable with an average decrease in variance of 14 percent for national means and 5 to 6 percent across all statistics. The specific domain efficiencies reported for cost model B ranged from 1.02 to 1.12. Table 1 summarizes the three-stage versus two-stage efficiencies for cost models A and B with the mobility rate set at .40. Tables including corresponding results for the other two mobility levels are presented in chapter V.

The gains in efficiency reported in table 1 may be slightly understated since the NAEP type stratification of PSUs reflected in our three-stage variance components model was limited by an "all-State" representation requirement. Also, the optimum NLS three-stage design calls for 500 or 600 PSUs while the NAEP sample had 116 primaries. With more primaries indicated, a deeper and more effective primary stratification scheme could be developed. It is unlikely, however, that the negligible to modest gains in precision projected in chapter V for the optimized NLS three-stage design would be enhanced enough to justify a major redesign effort.

Table 1. Efficiencies of the Optimum Three-stage Design Relative to the Optimum Two-stage Design

Domain	Number of Statistics	Relative efficiencies	
		Cost Model A	Cost Model B
National means	14	1.05	1.14
Subgroup means			
Whites	14	1.01	1.07
Blacks	14	1.01	1.06
Other races	14	1.02	1.07
Females	14	1.05	1.12
Males	14	1.02	1.08
Father-high school graduate	14	1.00	1.04
Father-less than high school	14	0.99	1.02
Father-more than high school	14	1.01	1.05
All domain means	112	1.01	1.06
Differences between domain and national means	112	1.00	1.04
All statistics	238	1.01	1.06
Optimum designs			
Two-stage			
Schools		1,263	1,010
Seniors per school		17	22
Total seniors		21,471	22,220
Three-stage			
PSU's		609	521
Schools per PSU		4	5
Seniors per school		9	9
Total seniors		21,924	23,445

II. THE TWO- AND THREE-STAGE VARIANCE COMPONENT MODELS

For variance components analysis, the NLS sample can be viewed as a nested structure with three levels. The first level is fixed and corresponds to the set of final strata used to stratify the schools in the sampling frame. The two sampling levels nested within the final strata are schools and students within schools. Since the first level is fixed and the last two are random, it can be shown that the variance of a statistic $\hat{\theta}$ for a two-stage design with n_1 schools and n_2 students per school can be written as

$$V2(\hat{\theta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2}$$

where σ_1^2 is the between school within final stratum variance component for statistic $\hat{\theta}$, and

σ_2^2 is the between student within school variance component for statistic $\hat{\theta}$.

The deeply stratified three-stage NLS design under consideration can also be viewed as a nested structure but with four levels. The first level is fixed and corresponds to the set of final strata used to stratify the primary units. The three sampling levels nested within primary strata are considered random and are respectively PSUs within strata, schools within PSUs, and students within schools. For the three-stage model (four levels with the first level fixed and the last three random), the variance of the statistic $\hat{\theta}$ with m_1-PSUs, m_2-schools per PSU,

and m_3-students per school is

$$V3(\hat{\theta}) = \frac{v_1^2}{m_1} + \frac{v_2^2}{m_1 m_2} + \frac{v_3^2}{m_1 m_2 m_3}$$

where v_1^2 is the between PSU within stratum variance component,
v_2^2 is the between school within PSU variance component, and
v_3^2 is the between student within school variance component.

In the next section, a cost-variance efficiency criterion for contrasting a three-stage NAEP-type design with a two-stage NLS-type design will be proposed. To estimate these efficiencies, NLS two-stage variance components based on the 14 change-year variables described in the paper by Moore and Shah [3] were utilized. For each variable, 17 sets of variance components had been estimated: one set associated with the national mean, eight sets associated with domain means, and eight sets associated with differences between domain means and the national mean. The specific domains or subpopulations considered were whites, blacks, other races, females, males; and three levels of father's education: less than high school education, high school graduate, and some education beyond high school.

Associated three-stage variance components were obtained from Year 05 (1974) of National Assessment where variance components had been estimated by combining 17-year-old in-school student results across 14 of the 16 group administered packages. NAEP components were available for 18 experimental exercises that had been repeated in from 3 to 9 of the group packages. For each of these 18 exercises, components had been estimated for the national

proportion answering correctly, for 20 domain proportions correct, and for the corresponding 20 differences between domain and national proportions. The 20 NAEP reporting groups included the 8 specific NLS domains mentioned above.

The basic concern here is not to compare the NLS 1972 cohort design with the NAEP Year 05 design, but rather to contrast the variance of a three-stage sample designed specifically for the NLS new cohort with the variance obtained from an optimized version of the 1972 cohort two-stage design. Several crucial assumptions were made when translating NLS two-stage components into three-stage components via NAEP component ratios. The first of these assumptions relates to the relative impact of unequal weighting on the two designs. While the Year 05 NAEP design exhibits greater student weight variability and a higher associated variance inflation factor than is observed in the 1972 NLS weights, we have assumed that a comparable level of unequal weighting could be achieved in an NLS three-stage design. This assumption is predicated on utilizing the same computerized national school frame used for the two-stage sample to approximate senior enrollments by PSU. With these accumulated enrollments available as PSU size measures, a three-stage PSU and school within PSU selection scheme could be designed to yield school selection probabilities proportional to senior enrollment. To oversample schools located in low income areas and those with high percentages of minorities, the enrollments of such schools could be doubled when the PSU accumulations were formed. Using these adjusted size measures to select primary units and to allocate schools to SES strata within PSUs would accomplish the two-to-one oversampling desired. Since one would expect school and student response rates to be similar for the two- and three-stage samples, it should be possible to

control all the design elements contributing to unequal weighting at comparable levels for the two designs; that is, inaccurate size measures, oversampling for specific subpopulations, and nonresponse at the school and student level should have comparable effects on the weights for both designs. The assumption of equal weight distributions allows one to claim that student within school variance components should be the same for both designs. This assumption also leads one to equate the total of all variance components above the student level in both designs. That is, if v_0^2 denotes the between primary stratum component for a three-stage sample and σ_0^2 the final school stratum component in the two-stage design, then we claim

$$\Sigma S \equiv (\sigma_0^2 + \sigma_1^2) = (v_0^2 + v_1^2 + v_2^2) \equiv VS.$$

The aggregate components ΣS and VS both represent a "schools within USA" variance. Finally, we have assumed that the partitioning of the total school variance component into three-stage primary strata, PSU, and school within PSU components will follow the same pattern across reporting domains as exhibited by the 1974 NAEP component ratios $R_0 = v_0^2/VS$, $R_1 = v_1^2/VS$, and $R_2 = v_2^2/VS$. In addition to reflecting the less than optimal NAEP "all-State" stratification scheme, these values of R are more comparable to NLS base-year statistics than to the NLS change-year variance components used here. Without the "all-state" requirement, a primary stratification scheme more favorable to the three-stage option could be designed, but the effect of such improvements on the total three-stage variance would probably be slight. The NLS change-year variance components differ from their base-year analogues primarily through a reduction in the between school and stratum components. We have assumed that the shift from base-year to change-year components will be

properly reflected by associated shifts in "school within USA" aggregates, ΣS, with the allocation fractions (Rs) for strata, PSU's, and schools remaining stable. Details of how the NAEP R ratios and NLS aggregate components are used to specify the relative efficiency of our competing designs are presented in the following section.

III. THE EFFICIENCY CRITERION TO COMPARE A THREE-STAGE VERSUS A TWO-STAGE SAMPLE DESIGN

The criterion for determining whether a two- or three-stage sampling design would be more cost efficient in selecting the new NLS cohort is developed in this section. The design selected should perform reasonably well over the entire range of statistics to be reported by NLS. Since many statistics are to be reported from a single sample, the criterion for choosing such optimum designs is not obvious. Moore, Shah, and Folsom [1] have provided an empirical solution for the parameters n_1 and n_2 which would maximize the average efficiency of a two-stage sample design subject to the fixed-cost constraint

$$C-C_0 = n_1C_1 + n_1n_2\,C_2 \quad . \tag{1}$$

Recalling that the two-stage variance of a statistic $\hat{\theta}_k$ takes the form

$$V2(k) = \sigma_1^2(k)/n_1 + \sigma_2^2(k)/n_1n_2$$

which has a minimum

$$MINV2(k) = \left(\sqrt{\sigma_1^2(k)C_1} + \sqrt{\sigma_2^2(k)C_2}\right)^2/(C-C_0)$$

subject to the linear cost model in (1), the efficiency of a design characterized by parameters $n_1(d)$ and $n_2(d)$ compared to the optimum design for statistic-k is

$$E_k(d) = MINV2(k)/\{\sigma_1^2(k)/n_1(d) + \sigma_2^2(k)/n_1(d)\ n_2(d)\} \quad . \tag{2}$$

The average efficiency criterion explored in [1] was simply

$$E(d) = \sum_{k=1}^{T} E_k(d)/T \qquad (3)$$

where the summation in (3) extended over various groupings of national proportions, domain proportions, and differences between domain and national percentages. A set of cost equivalent designs was established by specifying a range of $n_2(d)$ values and solving for the fixed cost value of $n_1(d)$ from equation (1). The 1972 cohort design with $n_2 = 18$ and $n_1 = 1{,}200$ had the largest average efficiency when the average in (3) extended over all three types of statistics mentioned above.

An alternative analytic solution for the two-stage optimum can be obtained by minimizing the average inefficiencies

$$L2(d) \equiv \sum_{k=1}^{T} E_k^{-1}(d)/T$$

$$= Z_1^2/n_1(d) + Z_2^2/n_1(d)\ n_2(d) \qquad (4)$$

where

$$Z_j^2 = \sum_{k=1}^{T} \{\sigma_j^2(k)/MINV2(k)\}/T \quad \text{for } j=1,\ 2\ .$$

The Z_j^2 components in (4) incorporate the two-stage variance components for all statistics k and the linear cost components C_1 and C_2 through MINV2 but are free of the design parameters $n_1(d)$ and $n_2(d)$. The values of n_1 and n_2 which minimize (4) subject to the cost constraint in (1) are therefore

$$n_2(\phi) = \sqrt{Z_2^2 \cdot C_1/Z_1^2 \cdot C_2}$$

and

$$n_1(\phi) = (C-C_0)\sqrt{Z_1^2/C_1} \Big/ \sum_{j=1}^{2} \sqrt{Z_j^2 \cdot C_j}\ .$$

Similarly, the optimum three-stage design can be obtained by minimizing average inefficiencies

$$L3(d) = \sum_{k=1}^{T} L3_k(d)/T$$

$$= W_1^2/m_1(d) + W_2^2/m_1(d)\, m_2(d) + W_3^2/m_1(d)\, m_2(d)\, m_3(d) \qquad (5)$$

where

$$W_j^2 = \sum_{k=1}^{T} \{v_j^2(k)/MINV3(k)\}/T \quad \text{for } j=1, 2, 3$$

and

$$MINV3(k) = \left\{ \sum_{j=1}^{3} \sqrt{v_j^2(k) \cdot \$_j} \right\}^2 \Big/ (\$-\$_0) \ .$$

The components \$ refer to a three-stage cost model of the form

$$(\$-\$_0) = m_1\$_1 + m_1m_2\$_2 + m_1m_2m_3\$_3 \qquad (6)$$

where m_1 denotes the number of sample PSUs, m_2 the number of sample schools per PSU, and m_3 the number of sample seniors per school. The same fixed value for variable costs that was specified for the two-stage design will also be used in the three-stage cost constraint; that is, $(C-C_0)$ from equation (1) equals $(\$-\$_0)$ from equation (6). The v2 three-stage variance components are as defined in chapter II. The values of m which minimize the average loss in (5) subject to the fixed cost in (6) are

$$m_3(\phi) = \sqrt{W_3^2 \cdot \$_2 / W_2^2 \cdot \$_3}$$

$$m_2(\phi) = \sqrt{W_2^2 \cdot \$_2 / W_1^2 \cdot \$_2}$$

and

$$m_1(\phi) = (\$-\$_0)\sqrt{W_1^2/\$_1} \Big/ \sum_{j=1}^{3} \sqrt{W_j^2 \cdot \$_j} \ .$$

Having obtained these optimum or "best-compromise" design parameters the three-stage versus two-stage efficiency factor $E_{3/2}(k)$ for statistic-k could be computed as

$$E_{3/2}(k) = \left\{ \frac{\sigma_1^2(k)/n_1(\phi) + \sigma_2^2(k)/n_1(\phi)n_2(\phi)}{v_1^2(k)/m_1(\phi) + v_2^2(k)/m_1(\phi)m_2(\phi) + v_3^2(k)/m_1(\phi)m_2(\phi)m_3(\phi)} \right\}.$$

Thus, the average efficiency of the best-compromise three-stage design relative to the optimum two-stage design over a particular class of statistics like the Black subgroup proportions is

$$E_{3/2}(B) = \sum_{k \epsilon B} E_{3/2}(k)/T(B) .$$

While the efficiency criterion based on best-compromise versions of the two- and three-stage samples is appealing, the amount of additional computations required was considered excessive. A shortcut method was therefore applied to approximate the average efficiencies. From the Moore and Shah report [3] and supporting data, averages of pertinent change-year variance component ratios were formed; namely,

$$\delta_0(k) \equiv \sigma_0^2(k)/[\sigma_0^2(k) + \sigma_1^2(k) + \sigma_2^2(k)]$$

$$\delta_1(k) \equiv \sigma_1^2(k)/[\sigma_0^2(k) + \sigma_1^2(k) + \sigma_2^2(k)]$$

and

$$\delta_2(k) \equiv \sigma_2^2(k)/[\sigma_0^2(k) + \sigma_1^2(k) + \sigma_2^2(k)] .$$

Table 2 presents average values of these component ratios by domain and type of statistic for NLS change variables.

Table 2
Average Values of the NLS Change-Year
Variance Component Ratios

Domain	δ_0	δ_1	δ_2	Number of Statistics
National means	.011	.017	.972	14
Subgroup means				
Whites	.011	.011	.978	14
Blacks	.009	.008	.983	14
Other races	.005	.006	.989	14
Females	.005	.009	.986	14
Males	.007	.007	.986	14
Father high school graduate	.006	.005	.989	14
Father less than high school	.008	.003	.989	14
Father more than high school	.005	.006	.989	14
All domain means	.008	.009	.983	112
Difference of domain and national means	.005	.009	.986	112
All statistics	.007	.008	.985	238

The optimum number of schools and students per school for domain-θ subject to a given two-stage cost constraint is approximated in terms of the $\delta_1(\theta)$ and $\delta_2(\theta)$ averages by

$$n_2(\alpha) = \sqrt{\delta_2(\theta) \cdot C_1 / \delta_1(\theta) \cdot C_2}$$

and

$$n_1(\alpha) = (C-C_0) \cdot \sqrt{\delta_1(\theta)/C_1} \Big/ \sum_{j=1}^{2} \sqrt{\delta_j(\theta) \cdot C_j} \qquad (7)$$

with the (α) being used to distinguish these approximate solutions from the previous $n(\phi)$ optima. Applying the two-stage cost components C_1 = \$368 and C_2 = \$180 from [3] along with the "all statistics" averages from table 2, the approximate solutions were $n_2(\alpha) = 17$ and $n_1(\alpha) = 1{,}263$. Compared with the empirical solution obtained by Moore and Shah where $n_2(\phi) = 18$ and $n_1(\phi) = 1{,}200$ the approximation proposed in equation (7) would seem adequate for making relative comparisons between designs. Similar component ratios were calculated for the three-stage design by combining 1974 NAEP 17-year-old averages with the NLS-δ ratios defined above. Recall that in chapter II the stage was set for utilizing NAEP component ratios of the form

$$R_0 \equiv v_0^2/\{v_0^2 + v_1^2 + v_2^2\}$$

$$R_1 \equiv v_1^2/\{v_0^2 + v_1^2 + v_2^2\}$$

and

$$R_2 \equiv v_2^2/\{v_0^2 + v_1^2 + v_2^2\}$$

which represented the primary stratum, PSU, and school within PSU fractions of the total school within USA component aggregate. For a particular domain or type of statistic-θ these ratios were used to partition the corresponding

NLS averages into primary stratum, PSU, and school within PSU fractions of the overall aggregate variance; that is,

$$\rho_0(\theta) \equiv R_0(\theta)\,[\delta_0(\theta) + \delta_1(\theta)]$$

is equivalent to the average

$$\rho_0(\theta) = \sum_{k\varepsilon\theta} \{v_0^2(k)/[v_0^2(k) + v_1^2(k) + v_2^2(k) + v_3^2(k)]\}/T(\theta)$$

when the three-stage components are obtained as

$$v_0^2(k) \equiv R_0(\theta)\,[\sigma_0^2(k) + \sigma_1^2(k)]$$

$$v_1^2(k) \equiv R_1(\theta)\,[\sigma_0^2(k) + \sigma_1^2(k)]$$

$$v_2^2(k) \equiv R_2(\theta)\,[\sigma_0^2(k) + \sigma_1^2(k)]$$

and

$$v_3^2(k) \equiv \sigma_2^2(k)$$

for domain-θ statistics. The other relevant fractions are approximated by

$$\rho_1(\theta) \equiv R_1(\theta)\,[\delta_0(\theta) + \delta_1(\theta)]$$

$$\rho_2(\theta) \equiv R_2(\theta)\,[\delta_0(\theta) + \delta_1(\theta)]$$

and

$$\rho_3(\theta) \equiv \delta_2(\theta)$$

representing, respectively, the PSU within primary stratum, school within PSU, and senior within school averages.

Table 3 provides values of the NAEP component ratios R_0, R_1, and R_2 for the various domain and statistic types presented in table 2. Extending

Table 3
Average Values of the Three-Stage Stratum, PSU, and School Variance Component Ratios*

Domain	R_0	R_1	R_2	Number of Statistics
National means	.167	.058	.775	18
Subgroup means				
Whites	.202	.103	.695	18
Blacks	.301	.094	.605	18
Other races	.453	.220	.327	18
Females	.201	.055	.744	18
Males	.175	.094	.731	18
Father less than high school	.529	.145	.326	36
Father high school graduate	.339	.146	.515	18
Father more than high school	.273	.153	.574	18
All domain means	.235	.108	.657	360†
Difference between domain and national means	.252	.104	.644	360
All statistics				

*The three-stage stratum, PSU, and school variance component ratios were calculated using the data from the Year 05 (1974) NAEP assessment of in-school 17-year-old students.

†Statistics for all 20 NAEP reporting domains were averaged.

the approximate solution presented in equation (7) to the three-stage design yields

$$m_3(\alpha) = \sqrt{\rho_3(\theta) \cdot \$_2/\rho_2(\theta) \cdot \$_3}$$

$$m_2(\alpha) = \sqrt{\rho_2(\theta) \cdot \$_1/\rho_1(\theta) \cdot \$_2}$$

and

$$m_3(\alpha) = (\$-\$_0) \cdot \sqrt{\rho_1(\theta)/\$_1} \; / \; \sum_{j=1}^{3} \sqrt{\rho_j(\theta) \cdot \$_j} \; .$$

The optimum values of the n and m parameters used to specify competing designs will be based on the "all statistics" averages from tables 2 and 3 and on six alternative sets of two- and three-stage cost components developed in the following chapter. The average relative efficiencies for these designs by domain will take the form

$$E_{3/2}(\theta) = \left\{ \frac{\delta_1(\theta)/n_1(A) + \delta_2(\theta)/n_1(A)\, n_2(A)}{\rho_1(\theta)/m_1(A) + \rho_2(\theta)/m_1(A)\, m_2(A) + \rho_3(\theta)/m_1(A)\, m_2(A)\, m_3(A)} \right\}$$

with the n(A) and m(A) parameters indicating "all statistics" optima.

IV. COST MODELS

The cost model for the three-stage sample design efficiency analysis includes the total variable costs incurred in conducting a base-year survey and five follow-up surveys. Modeling of the distance traveled by interviewers in the course of an NLS follow-up survey is described in part A. In part B the costs of conducting an NLS follow-up survey using a three-stage sample design are described. It was assumed that the costs not associated with the number of PSUs, schools, or students selected would be about the same for either a two- or three-stage design. In part C the costs of a NLS base-year survey are estimated assuming a three-stage sample design. In addition, the cost parameters for a linear cost model associated with the base-year and five follow-up surveys are determined. An alternative procedure for allocating costs and the resulting two- and three-stage cost models are discussed in part D. Cost models assuming different proportions of students moving from the two-county area in which their school is located are discussed in part E.

A. Distance Traveled Model

In the study "Choosing the 'optimum' Design" for an NLS Cohort", [3], a distance-traveled model was assumed that gave an indication of a cost reduction that might be obtained due to clustering sample students in groups of 9-24 per school. In a three-stage model, the clustering principle is extended to PSUs. Ideally the three-stage design should result in the schools being clustered in selected areas averaging two counties in size so that more students would be concentrated in the two county area than otherwise would be if a two-stage design was used to select sample students. Thus, with a

three-stage design, interviewers should spend less time traveling and locating the nonrespondents than with a two-stage design. As with the two-stage sample design, the cost reduction is dampened by the fact that the population is quite mobile. The distance model should also reflect interviewer's salary as well as travel costs.

The average distance between k points, assumed to be equally spaced, within an area A is $\bar{D}=\sqrt{\frac{A}{k}}$. The total distance traveled to visit all k points once is $D=\sqrt{kA}$. For the follow-up study, only the nonrespondents to the mail questionnaire will be personally interviewed. For the three-stage model, we will assume $D=D_1 + D_2$ where

D = the total distance traveled,

D_1 = the distance traveled to visit all nonrespondents who still live within the initial two county PSU of the school location, and

D_2 = the distance traveled to visit all nonrespondents who have moved from the PSU of the school location.

A working staff of I interviewers, each of whom visit m_1/I PSU locations is assumed where m_1 depicts the total number of sample PSUs. Within each PSU area, the interviewer will visit $(1-r)(1-\mu)\ m_2\ m_3$ nonrespondents who are still living in the two county PSU where

m_2 = number of sample schools per PSU,

m_3 = number of sample students per school,

r = proportion of sample students who will return a completed questionnaire by mail, and

μ = proportion of sample students who have moved out of the PSU in which their school is located.

It is assumed that one interviewer can interview all nonrespondents within a PSU. A three-stage design similar to the NAEP sample design has been assumed for the

cost modeling discussion. It is expected there will be approximately 15-17 large self-representing PSUs like the New York City SMSA which would require more than one interviewer. Such large units would probably be divided into smaller administrative units, one or more of which could be handled by a single interviewer. For cost modeling purposes these administrative units will be viewed as separate PSUs.

If we assume that the nonmoving students are equally spaced throughout the two county PSU of average area $\frac{2A}{3,141}$ square miles, then the distance traveled within the PSU is given by

$$\sqrt{(1-r)(1-\mu)\cdot m_2 \cdot m_3 \frac{2A}{3,141}} \quad .$$

Since there are m_1 PSUs, the total distance traveled to interview all nonrespondents who are still within the PSU area of their school location is given by

$$D_1 = m_1 \sqrt{(1-r)(1-\mu)\cdot m_2 \cdot m_3 \cdot \frac{2A}{3,141}} \quad .$$

In addition, each interviewer would be required to visit a total of $(1-r)\ \mu m_1 \cdot m_2 \cdot m_3 / I$ nonrespondents who have moved from their initial PSUs. The movers are also assumed to be equally spaced throughout the interviewer's work area, of average size, A/I. Then, D_2 in the model can be expressed by

$$D_2 = I \cdot \sqrt{\frac{(1-r)\,\mu m_1 \cdot m_2 \cdot m_3 \frac{A}{I}}{I}} \quad .$$

The total distance traveled in miles, under the assumptions of our model, can be written as $D = D_1 + D_2$ or

$$D = m_1 \sqrt{(1-r)(1-\mu) m_2 m_3 \cdot \frac{2A}{3,141}} + \sqrt{(1-r)\mu \cdot m_1 m_2 m_3 \cdot A} \quad .$$

If we assume $r = .60$, $\mu = .40$ and $A = 3,615,122$, D can be expressed as

$$D = 23.504\ m_1 \sqrt{m_2 m_3} + 760,539 \sqrt{m_1 m_2 m_3} \qquad \text{or}$$

$$D = \sqrt{m_1 m_2 m_3}\ (23.504 \sqrt{m_1} + 760,539).$$

The above distance model indicates that the total distance traveled by interviewers depends only on the total number of students to be selected ($m_1m_2m_3$) and the number of PSUs (m_1) to be selected. The distance traveled does not depend on the number of schools selected within a PSU. This result appears to be reasonable since it is anticipated that contact with the sample schools will involve only mail and telephone procedures. Interviewers are not expected to personally contact the selected schools. If we then fix $m_1m_2m_3$ equal to approximately 21,600 (the approximate total number of pupils assumed for a two-stage design), and allow m_1 to vary, we can rewrite the distance equation as

$$D = 3{,}454\sqrt{m_1} + 111{,}776.$$

The distance-traveled model is difficult to work with in this form. However, we can represent the model quite accurately by the linear function

$$D = 89\, m_1 + 144{,}298$$

within the range of m_1 in which we are interested, $200 \leq m_1 \leq 600$. Outside this range, the nonlinear distance equation indicates a slight curvature and hence, the above linear model is not as good an approximation.

A linear distance-traveled model was estimted using a two-stage sample design [3] and is given here by

$$D_{2\text{-stage}} = 65n_1 + 182{,}000$$

where n_1 is the number of schools to be selected and is between 600 and 1800. Outside of this range, the above linear two-stage distance model does not accurately describe the distance traveled. The number of pupils to be selected in both distance traveled models was assumed to be 21,600.

If we assume approximately 400 PSUs will be selected using a three-stage design and approximately 1200 schools will be selected using a two-stage design, then

$$D_{3\text{-stage}} = 89(400) + 144{,}298 = 179{,}898 \text{ miles},$$

and

$$D_{2\text{-stage}} = 65(1200) + 182{,}000 = 260{,}000 \text{ miles}.$$

We would expect a savings of 80,102 miles if a three-stage design is used. Since the direct charges are proportional to the distance traveled, we would expect a savings in direct charges for the three-stage design of approximately 30.8 percent $\left(\frac{80,102}{260,000} \times 100\right)$. Correspondingly, we would expect interviewer salary and mileage expenses to be 30.8 percent lower using the three-stage design than for a comparable number of students selected using the two-stage design.

The distance-traveled model can also be used to determine the proportion of interviewer travel and salary expenses that are PSU-related. We would expect $\left(\frac{89m_1}{89m_1 + 144,298}\right)$ or 19.8 percent of the total cost to be attributed to the PSU cost component while $\left(\frac{144,298}{89m_1 + 144,298}\right)$ or 80.2 percent to be related to student costs.

B. Cost of a Single Follow-Up Survey Using a Three-Stage Sample Design

A linear cost model of the form

$$\$ = \$_o + m_1\$_1 + m_1m_2\$_2 + m_1m_2m_3\$_3$$

will be used to model the cost of a follow-up survey where a three-stage design is used to select the students. The terms in the model are defined below as:

$\$$ = total cost of the follow-up survey,

$\$_o$ = fixed cost which does not depend on the sample design nor the number of students selected,

$\$_1$ = marginal cost of adding an additional PSU,

$\$_2$ = marginal cost of adding an additional school to the sample within a PSU,

$\$_3$ = marginal cost of adding an additional student to the sample within a school,

m_1 = number of sample PSUs,

m_2 = number of sample schools per PSU, and

m_3 = number of sample students per school.

The cost modeling has been based on costs estimated in the RTI report, "Choosing the 'Optimum' Design for an NLS Cohort", [3]. The modeling in that report was based on estimates provided by RTI's second follow-up proposal.

For the three-stage sample design under consideration, it is necessary to isolate the variable cost portion of a follow-up survey and to estimate $\$_1$, $\$_2$, and $\$_3$, the vital cost parameters of the model. Ten major tasks have been outlined in the above mentioned report [3]. They are

Task 1 - general project direction,

Task 2 - implementation of tracing procedures,

Task 3 - procurement, distribution, and receipt control of survey questionnaires,

Task 4 - development of critical data base,

Task 5 - computer software support, further development of the automated survey system,

Task 6 - reliability and validity studies,

Task 7 - field interviews of nonrespondents,

Task 8 - analysis of data,

Task 9 - preparation of special interest reports, and

Task 10 - design, preparation, and field test for an additional follow-up survey.

Tasks 1, 4, 5, 6, 8, 9, and 10 are assumed to involve only fixed costs. These fixed costs are assumed to be the same regardless of whether a two- or three-stage sample design is used. The variable portion of the survey cost is assumed to be associated with some of the costs necessary to perform tasks 2, 3, and 7.

Virtually all of the direct charges estimated for tasks 2, 3, and 7 can be regarded as variable costs. Exceptions are some of the reproduction costs from tasks 2 and 7, and the central staff travel costs from task 3. The majority of the costs are considered to occur in proportion to the total number of students in the sample, which can be assumed to be independent of whether a two- or three-stage design is used. Note, however, the distance model indicates a 30.8% savings in travel related cost for a three-stage sample design. This saving may allow more students to be selected. The interviewer salaries and mileage expenses are assumed to be proportional to the distance traveled. Based on the distance-traveled model, the interviewers' salary and mileage expenses are assumed to be dependent on the number of sample PSUs (m_1) and the total number of sample students ($m_1 m_2 m_3$) but are not related to the number of sample schools. In addition, the total interviewer salaries and mileage expenses are assumed to be 30.8% less using the three-stage sample design than the same expenses estimated for the two-stage sample design. Based on the distance-traveled model in part A, 19.8% of the interviewers' salaries and mileage expenses are considered PSU-related while 80.2% are student-related. Table 4 shows the dollar costs for direct charges, by cost item for PSU-, school-, and student-related cost categories. Based on the cost modeling done for the two-stage design [3], we are assuming that all of the direct labor charges are student-related and thus for both the two-stage cost model and the three-stage cost model that all of the direct labor costs have been allocated to the student component.

The cost parameters, $\$_1$, $\$_2$, and $\$_3$ (relating to the PSU-, school-, and student-related variable costs) were estimated based on table 4 and the report "Choosing an 'Optimum' Design for an NLS Cohort" [3]. The calculations are shown in table 5. Direct labor was converted to dollars using the same cost factors

given in [3]. The estimated direct charges are given in table 4. Overhead or indirect costs were included using the same RTI proposal rates for indirect technical expense and overall management and service expense as used for the two-stage cost model [3]. An average fixed fee of five percent was also included to arrive at the total estimated variable costs of $26,030 for PSU-related costs and $702,267 for student-related costs. Since only interviewer salaries and mileage expenses are apportioned to the PSU- and student-related costs, the costs for the follow-up survey are not related to the number of sample schools within a PSU. The distance-traveled model was used to determine the total cost and the proportion to be allocated to the PSU and student cost. Dividing the total PSU-cost by 400 and the total student cost by 21,600, the estimated cost parameters for the follow-up survey of $\$_1$ = $65.08, and $\$_3$ = $32.51 are obtained as shown in Table 5. The school cost component $\$_2$ for the follow-up survey is assumed to be zero.

C. Costs of All Surveys in An NLS Cohort

In order to optimize the sample design for all surveys of an NLS cohort, the cost parameters in table 5 were assumed representative for each of five follow-up surveys. The base-year cost parameters were developed in an earlier report [1] "Efficiency Study of NLS Base-Year Design", based on a two-stage sample design. Tables 1, 2, and 3 of the report [1] have been included in appendix B as tables B-1, B-2, B-3. The school related costs are considered to be independent of either a two- or three-stage sample design. Thus, the school- and student-related costs of the base year are related only to the total number of schools and students selected and are not dependent on the type of sampling design. The base-year cost parameters, $\$_2$, school-related, and $\$_3$, student-related should be the same as in RTI

Table 4
Estimated Direct Charges Allocated** to PSU, School, and Student Costs

Item	PSU-Cost ($\$_1$) Dollars	Student-Cost ($\$_3$) Dollars
Printing subcontract		86,940
Other reproductions		1,750
Materials		4,150
Communications and shipping		88,730
Services		4,225
Temporary labor		104,089
Staff travel		40,500*
Interviewer salaries	12,468*	50,504*
Interviewer mileage	9,865	39,959
Total Direct Charges	22,333	420,847

*The total for interviewers' salaries and expenses was estimated at 69.2% of the total for the 2-stage design (69.2% of 91,000 is 62,972 and 69.2% of 72,000 is 49,824). These expenses were then allocated 19.8% for PSU-cost and 80.2% for student-cost.

**All direct charges are considered to be either PSU- or student-related. School-related costs are assumed to be zero.

Table 5
Calculation of Estimated Variable Cost Parameters,
$\$_1$, $\$_2$, and $\$_3$ for an NLS Follow-Up Survey
Using a Three-Stage Sampling Design.*

Item	PSU-Cost ($\$_1$) Dollars	Student-Cost ($\$_3$) Dollars
Direct labor		
Survey specialists		55,080
Research assistants		34,310
Survey clerks		14,438
Total		103,828
Indirect technical expense[1/]		77,871
Direct charges	22,333	420,847
Overall management and service expense[2/]	2,457	66,280
Total estimated costs	24,790	668,826
Fixed fee	1,240	33,441
Total estimated cost plus fixed fee	26,030	702,267
Number of PSUs and students assumed selected	400	21,600
Estimated cost parameters[4/]	65.08	32.51

[1/] Computed as 75 percent of direct labor.

[2/] Computed as 11 percent of direct labor, indirect technical expense, and direct charges.

[3/] Calculated as five percent of total estimated costs.

[4/] Total estimated cost plus fixed fee divided by the number assumed selected.

*The school-related cost is assumed to be zero.

report [3] with $\$_2 = \$130.38$ and $\$_3 = \$11.28$. The base-year cost of adding a PSU is considered to be zero since the base-year data collection operation is by mail.

The total cost model

$$\$ - \$_0 = m_1\$_1 + m_1m_2\$_2 + m_1m_2m_3\$_3$$

can be used to represent the cost of an NLS base-year survey and five follow-up surveys for a three-stage design. Each cost component in the model equals the respective base-year component plus five times the associated follow-up cost component. Thus, the PSU cost parameter $\$_1 = \$0.0 + 5(65.08) = \$325.40$, the school cost parameter $\$_2 = \$130.38 + 5(0.0) = \$130.38$, and the student cost parameter $\$_3 = \$11.28 + 5(32.51) = \$173.83$. Substituting the above values for their corresponding cost parameters yields the following cost model for a three-stage sample design

$$\$ - \$_0 = m_1(\$325.40) + m_1m_2(\$130.38) + m_1m_2m_3\ (\$173.83),$$

where $\$ - \$_0$ is fixed at the same level as in the two-stage cost model developed by Moore and Shah [3] and is equal to \$4,329,600.

D. An Alternative Procedure In Determining Both The Two- and Three-Stage Cost Models.

The cost modeling in parts B and C of this section was patterned along the lines used for the two-stage cost modeling done in the reports "Choosing the 'Optimum' Design for an NLS Cohort" [3] and "Efficiency Study of NLS Base-Year Design" [1]. To examine the sensitivity of these cost components to alternative assumptions, one could first presume that the survey specialists labor and staff travel expenses are related to the number of interviewers trained and supervised. Since the number of interviewers required is dictated by the distance traveled, one could argue that the distance traveled model should be used to apportion these costs as PSU-, school-, and student-related. For the two-stage design, Moore and Shah [3] apportioned 30 percent of the travel

and salary costs as school-related and 70 percent as student-related. Thus, the survey specialists' labor and staff travel expenses have been apportioned as 30 percent school-related and 70 percent student-related. Applying the same argument to the three-stage design the above two cost items were judged to be 19.8 percent PSU-related and 80.2 percent student-related with a reduction of 30.2 percent in total costs from the corresponding two-stage costs.

The estimted direct charges for the follow-up survey are shown in appendix table B-4, part B for the two-stage design, and part A for the three-stage design. Table B-5 then gives the total estimated school- and student-related cost for the two-stage model while table B-6 gives the total estimted PSU-, school- and student-related costs for the three-stage model. These tables have been included in appendix B for ease of readability since they are similar in format to tables 4 and 5. Assuming a total of approximately 21,600 students and 1,200 schools will be selected, the two-stage cost parameters for the follow-up survey are given by C_1 (school cost component) = \$87.38 and C_2 (student cost component) = \$31.57. In estimating the three-stage cost model, it was assumed approximately 21,600 students and 400 PSUs will be selected. The cost components are then given by $\$_1$ (PSU cost component) = \$119.72, $\$_2$ (school cost component) = \$0.0 and $\$_3$ (student cost component) = \$29.23.

Since the base-year costs are not affected by the distance traveled to interview nonrespondents, the base-year cost components will be the same as those given in part C of this section. The total cost model for the two designs is then given by adding the base year cost component to five times the follow-up survey cost component. For the two designs, the calculations are:

(1) Two-Stage Design

$$C_1 = \$130.38 + 5(\$87.38) = \$567.28$$

C_2 = \$11.28 + 5(\$31.57) = \$169.13, for a linear cost model of

$C-C_0$ = \$567 n_1 + \$169n_1n_2.

(2) Three-Stage Design

$\$_1$ = \$0.0 + 5(\$119.72) = \$598.60

$\$_2$ = \$130.38 + 5(0.0) = \$130.38

$\$_3$ = \$11.28 + 5(\$29.23) = \$157.53, for a linear cost model of

$\$-\$_0$ = \$599 m_1 + \$130m_1m_2 + \$157 $m_1m_2m_3$.

The total variable cost $C-C_0$ and $\$-\$_0$ are both assumed to be equal to \$4,329,600.

E. Cost Models For Two Proportions Of Students Moving and Two Methods Of Allocating Costs.

In determining the total distance traveled by interviewers, it was assumed that 40 percent of the students selected from a PSU would move out of the PSU by the time of the first follow-up survey. An analysis of the mobility of the sample NLS students from their base-year county location to their first follow-up and second follow-up location is planned and will be presented in another report.

In order to determine the sensitivity of our results to this parameter, two additional proportions of students moving were considered, namely 30 percent and 50 percent. A linear distance traveled model was determined for each of the two additional values of μ for both the two-stage and three-stage designs using the same reasoning as in part A of this section.

For $\mu = .50$ the linear distance traveled models are

$$D_{2\text{-stage}} = 192{,}456 + 61n_1,$$

and

$$D_{3\text{-stage}} = 155{,}894 + 77m_1,$$

where n_1 is the number of schools selected and is between 600 and 1800, and m_1 is the number of PSUs selected and is between 200 and 700. The average number of schools selected, $n_1 = 1200$, and the average number of PSUs selected, $m_1 = 450$, were used in calculating the average distance to be traveled for both the two and three-stage designs. The average distance traveled in miles using the two types of designs are

$$D_{2\text{-stage}} = 265{,}656 \text{ miles},$$

and

$$D_{3\text{-stage}} = 190{,}544 \text{ miles}.$$

One would then expect the interviewers' salaries and mileage expenses for the new two-stage model to be 102.2[1/] percent of the corresponding expenses listed in table 1 of the report, "Choosing the 'Optimum' Design for an NLS Cohort" [3]. This table is included in the appendix as table B-7, and tables 2 and 3 of reference [3] are also included as tables B-8 and B-9 in the appendix. For the three-stage design these interviewer costs would be 73.3 percent of the table B-7 amounts.

For $\mu = .30$ the linear distance traveled models are

$$D_{2\text{-stage}} = 170{,}964 + 67n_1,$$

1/ $\frac{265{,}656}{260{,}000} \times 100 = 102.2$

and

$$D_{3\text{-stage}} = 133,395 + 91m_1$$

over the same range of n_1 and m_1 as for μ=.60. If n_1=1200 and m_1=450, the average distance traveled is

$$D_{2\text{-stage}} = 251,364 \text{ miles}$$

and

$$D_{3\text{-stage}} = 174,345 \text{ miles.}$$

The interviewers' salaries and mileage expenses for the two-stage model are 96.7 percent of those listed in table B-7 in the appendix and 67.1 percent for the three-stage model.

In determining a linear cost model, two procedures are considered. In procedure A, only the interviewers' salaries and mileage expenses are considered dependent on the distance traveled. All other variable costs are considered student related. This is the procedure used in determining the cost model in parts B and C of this section. In procedure B, staff travel expenses and labor for the survey specialists are also considered to be related to the number of interviewers required which is in turn related to the distance traveled. With this assumption corresponding reductions in staff travel and survey specialists labor were made; and these line items were then allocated partially to schools and PSUs. This is the procedure used for specifying the alternative cost model in part D of this section.

The linear cost models determined for the two types of costing procedures and for the two values of μ are:

1. Procedure A, μ = .30

   $C\text{-}C_{0(2\text{-stage})} = \$374n_1 + \$178n_1n_2$,

   $\$\text{-}\$_{0(3\text{-stage})} = \$333m_1 + \$130m_1m_2 + \$172m_1m_2m_3$;

2. Procedure A, μ = .50

   $C\text{-}C_{0(2\text{-stage})} = \$354n_1 + \$182n_1n_2$,

   $\$\text{-}\$_{0(3\text{-stage})} = \$282m_1 + \$130m_1m_2 + \$176m_1m_2m_3$;

3. Procedure B, $\mu = .30$

$$C\text{-}C_{0(2\text{-stage})} = \$580n_1 + \$166n_1n_2,$$

$$\$\text{-}\$_{0(3\text{-stage})} = \$612m_1 + \$130m_1m_2 + \$154m_1m_2m_3;$$

4. Procedure B, $\mu = .50$

$$C\text{-}C_{0(2\text{-stage})} = \$541n_1 + \$172n_1n_2, \text{ and}$$

$$\$\text{-}\$_{0(3\text{-stage})} = \$518m_1 + \$130m_1m_2 + 161m_1m_2m_3.$$

The total variable costs, $C\text{-}C_0$ and $\$\text{-}\$_0$, are assumed to be equal to $4,329,600 for all of the above cost models.

Table 6
Optimum Number of Schools and Students Per School Selected For Each Domain Variable Subject To The Cost Model $C-C_0=\$368n_1 + \$180n_1n_2$.

Domain	$[n_1]_{opt}$ Schools	$[n_2]_{opt}$ Students	Total Number Of Students
National means	1,844	11	20,284
Subgroup means			
Whites	1,599	13	20,787
Blacks	1,333	16	21,328
Other Races	1,263	17	21,471
Females	1,411	15	21,165
Males	1,333	16	21,328
Father high school graduate	1,091	20	21,820
Father less than high school	857	26	22,282
Father more than high school	1,200	18	21,600
All domain means	1,411	15	21,165
Differences of domain and national means	1,200	18	21,600
All Statistics	1,263	17	21,471

Table 7
Optimum Number of PSUs, Schools per PSU, and Students Per School Selected For Each Domain Variable Subject To The Cost Model
$\$-\$_o = \$325m_1 + \$130m_1m_2 + \$174m_1m_2m_3$

Domain	$[m_1]_{opt}$ PSUs	$[m_2]_{opt}$ Schools	$[m_3]_{opt}$ Students	Total Number of Students
National means	588	6	6	21,168
Subgroup means				
Whites	757	4	7	21,196
Blacks	675	4	8	21,600
Other races	793	2	14	22,204
Females	458	6	8	21,984
Males	675	4	8	21,600
Father high school graduate	671	3	11	22,143
Father less than high school	793	2	14	22,204
Father more than high school	671	3	11	22,143
All domain means	675	4	8	21,600
Differences of domain and national means	609	4	9	21,924
All Statistics	609	4	9	21,924

Table 8
Optimum Number Of Schools And Students Per School
Selected For Each Domain Variable Subject To
The Cost Model, $C-C_0 = \$567n_1 + \$169n_1n_2$

Domain	$[n_1]_{opt}$ Schools	$[n_2]_{opt}$ Students	Total Number of Students
National means	1,476	14	20,664
Subgroup Means			
Whites	1,259	17	21,403
Blacks	1,097	20	21,940
Other races	1,010	22	22,220
Females	1,146	19	21,774
Males	1,097	20	21,940
Father high school graduate	873	26	22,698
Father less than high school	705	33	23,265
Father more than high school	937	24	22,488
All domain means	1,146	19	21,774
Differences of domain and national means	937	24	22,488
All Statistics	1,010	22	22,220

Table 9
Optimum Number of PSUs, Schools per PSU, and Students
Per School Selected For Each Domain Variable Subject
To The Cost Model
$\$-\$_o = \$599m_1 + \$130m_1m_2 + \$157m_1m_2m_3$

Domain	$[m_1]_{opt}$ PSUs	$[m_2]_{opt}$ Schools	$[m_3]_{opt}$ Students	Total Number of Students
National means	472	8	6	22,656
Subgroup means				
Whites	543	6	7	22,806
Blacks	521	5	9	23,445
Other races	538	3	15	24,310
Females	335	8	9	24,120
Males	439	6	9	23,706
Father high school graduate	500	4	12	24,000
Father less than high school	538	3	15	24,210
Father more than high school	539	4	11	23,716
All domain means	521	5	9	23,445
Differences of domain and national means	476	5	10	23,800
All statistics	521	5	9	23,445

Table 10
Relative Efficiencies of the "Best" Three-Stage Design Relative to the "Best" Two-Stage Design Assuming the Optimum Design Parameters for the "All Statistics" Group and Cost Procedure A

Domain	Relative efficiencies, using cost procedure A		
	μ=.30	μ=.40	μ=.50
National means	1.05	1.05	1.05
Subgroup means			
Whites	1.01	1.01	1.01
Blacks	1.01	1.01	1.01
Other races	1.02	1.02	1.02
Females	1.05	1.05	1.05
Males	1.02	1.02	1.02
Father high school graduate	1.00	1.00	1.00
Father less than high school	0.99	0.99	0.99
Father more than high school	1.01	1.01	1.01
All domain means	1.01	1.01	1.01
Differences of domain and national means	1.00	1.00	1.00
All statistics	1.01	1.01	1.01
The "optimum" designs are			
Two-stage			
Schools	1,273	1,263	1,256
Students per school	17	17	17
Total students	21,641	21,471	21,352
Three-stage			
PSUs	615	609	607
Schools per PSU	4	4	4
Students per school	9	9	9
Total students	22,140	21,924	21,852

Table 11
Relative Efficiencies of the "Best" Three-Stage Design Relative to the "Best" Two-Stage Design Assuming the Optimum Design Parameters for the "All Statistics" Group and Cost Procedure B

Domain	Relative Efficiencies, using cost procedure B		
	μ=.30	μ=.40	μ=.50
National means	1.14	1.14	1.13
Subgroup means			
Whites	1.08	1.07	1.07
Blacks	1.07	1.06	1.06
Other races	1.07	1.07	1.06
Females	1.12	1.12	1.11
Males	1.08	1.08	1.07
Father high school graduate	1.04	1.04	1.04
Father less than high school	1.02	1.02	1.02
Father more than high school	1.06	1.05	1.05
All domain means	1.07	1.06	1.06
Differences of domain and national means	1.05	1.04	1.04
All statistics	1.06	1.06	1.05
The "optimum" designs are			
Two-Stage			
Schools	1,023	1,010	1,043
Students per school	22	22	21
Total students	22,506	22,220	21,903
Three-stage			
PSU	529	521	515
Schools per PSU	5	5	5
Students per school	9	9	9
Total students	23,805	23,445	23,175

V. RESULTS OF COMPARING THE RELATIVE EFFICIENCIES OF A THREE- VERSUS A TWO-STAGE SAMPLE DESIGN

The methodology for comparing the relative efficiencies of a three- versus a two-stage sample design was developed in section III. The "best" optimum design parameters for both the two- and three-stage designs are presented in this section for each of the six cost models assumed in section IV. The six cost models are for three student mover proportions and two cost allocation procedures. The relative efficiencies, $E_{3/2}(\theta)$ are calculated for all 12 groups of statistics using the "best" design for the "all statistics" average. Variance component ratios for the NLS change-year variables given in table 2 section III, and NAEP variance component ratios R_1 and R_2, given in table 3 section III, were used to determine the optimum design parameters subject to fixed cost constraints as specified by the six paris of cost models derived in section IV. The optimum design parameters are presented in tables 6-9 for the two cost allocation procedures A and B assuming forty percent of the students will move from their initial PSU location. Cost allocation procedure A assumes that only a portion of the direct charges were related to interviewer travel. For cost alloction B, salaries and travel expenses of survey specialists were also partitioned among the PSUs, schools, and student components accoring to fractions based on the distance-traveled model. Tables 6 and 7 are based on allocation procedure A; the corresponding results for procedure B are shown in tables 8 and 9. The number of students selected per school and the number of schools selected from each PSU for the "best" three-stage designs, and the number of students selected per school for the "best" two-stage design did not change when the proportion of student movers varied from 30 percent

to 50 percent. The number of PSUs selected for the "best" three-stage design and the number of schools selected for the "best" two-stage design varied only slightly as the mover proportion varied within the above range. Differences in "optimum" design parameter values between the two cost procedures were more pronounced. For this reason, the complete set of optimum design parameters representing all 12 groups of statistics are presented for only one of the three mobility rates, namely $\mu = .4$. Optimum design parameters for the other two mobility rates ($\mu = .3$ and .5) are similar within cost allocation procedures. The "best" design parameter values for the "all statistics" category are presented in tables 10 and 11 for both the two- and three-stage designs for all six cost models.

The relative efficiencies, $E_{3/2}(\theta)$, of the "best" three-stage versus "best" two-stage design are presented in table 10 assuming cost procedure A. Table 11 presents the relative efficiencies assuming cost procedure B. The "optimum" designs are best in terms of "all statistics" averages. The relative efficiencies are not affected by the percent of students moving at least in the range of 30 to 50 percent movers. However, as the percent of movers increases, the total number of students selected decreases. The method of allocating the costs has more of an affect on the relative efficiencies than has the proportion moving. At first glance this result appears surprising. However, only the nonrespondents to the mail questionnaire will be interviewed personally. This dampens somewhat the effect that the number of movers has on the cost models and thus the effect on the relative efficiencies. The greater the mobility the less the advantage of the three-stage design as compared to the two-stage design.

Values of $E_{3/2}(\theta)$ less than 1.0 indicate that a two-stage design would be more efficient than a three-stage design subject to the cost constraints

and the manner in which the variance components were estimated. A value of 1.0 would indicate that either type of design should result in about the same total variance estimate for the given cost. The findings indicate that decreasing the student cost component results in relative efficiencies greater than 1.0 in spite of an increase in the PSU cost component and two-stage school cost component. The results indicate that procedure B which allocates a smaller proportion of costs to the student component and a larger portion to the PSU or school components leads to higher three-stage efficiencies than procedure A.

If cost procedure A is assumed to represent the true cost picture, then there is very little difference in total variance between the best two- and three-stage designs. The variance for the national means would be about 5 percent smaller with a three-stage design than with a two-stage design. However, averaged over the "all statistics" category only a one percent difference in total variance is observed between the two designs.

If cost procedure B is assumed, then the "best" three-stage design should result in a reduction in total variance of between 2 and 12 percent over the eight specific domains. For the "all statistics" category the percent reduction is 5 or 6 percent. For comparison sake, one can show that this design (521, 5, 9) represents as great an improvement over the (1010, 22) design as that design is an improvement over the (2177, 9) design considered by Moore and Shah [3].

The optimum number of PSUs derived for the three-stage design is larger than was expected, but it is not outside the range for which the cost model is valid. The 1974 NAEP design had, for example, only 116 PSUs selected from a primary frame containing 1,148 units. While the NAEP

analogy appears weak at this point, it should be pointed out that the 17 self-representing NAEP PSUs contained approximately 30 percent of the 1,198 17-year-old schools. The self-representing schools were partitioned into 60 administrative units with an average of six schools each. If 30 percent of the 2,605 schools specified for the three-stage model B design with $\mu = .4$ were allocated to 17 self-representing SMSAs, then 365 PSUs with five sample schools each would remain to be selected from the 1,131 non-self-representing units in the frame. Considering that the NAEP design was optimized for a single group package with a total planned sample size of around 2,600 responses, it is not surprising that the proposed NLS three-stage sample designed for approximately 22,000 responses would require many more primary units. Thds selection of more PSUs would permit a more deeply stratified sample which should reduce the variance for the "best" three-stage design. The resulting increase in relative efficiency is not expected to be substantial, however.

VI CONCLUSIONS

Relative efficiencies of a three-stage versus a two-stage design indicate a reduction in variance for the "best" three-stage design when survey specialist salaries and travel expenses in addition to other direct charges are allocated to the cost components using a distance-model for interviewer travel. When only the direct charges are allocated using the distance-traveled model, the projected increase in precision for a three-stage design is negligible.

In order to properly evaluate the advantage or disadvantage in using a three-stage design as opposed to a two-stage design, it would be well to review the assumptions made during the analysis. While cost modeling described in section IV was done in a careful manner considering the cost data available, separation of the variable costs involve subjective judgments, and as such, the allocation might be questioned. The two methods of allocating costs attempted to investigate how sensitive the relative efficiencies were to variation in the cost components.

Another assumption which had an impact on cost component estimates involved the proportion of students who were likely to move out of their initial PSU location. The more mobile the student population, the fewer students that can be selected for the same cost. As mentioned in section IV, RTI is in the process of determining what proportion of students have moved out of their base-year county by the time of the first follow-up and by the time of the second follow-up. These results will be presented in a later report. To allow for uncertainty regarding this parameter, modeling was done assuming the proportion of movers would be between 30 and 50%. It is interesting to note that the relative efficiencies were extremely

stable over the above range of mover proportions. In addition, the design parameters varied only slightly for the different mover proportions. This result indicates that under the distance traveled model of section IV the "best" design is not extremely sensitive to the mobility of sample students at least within the 30 to 50% range.

In the methodology section, section III of this report, the NLS variance components were partioned into equavalent three-stage variance components using ratios of NAEP Year 05 variance components. The variance component estimates for both the NLS and NAEP surveys are influenced by the corresponding stratification procedures. The NLS was a deeply stratified sample while the NAEP stratification of PSUs was not nearly as extensive. NAEP selected only 116 PSUs from a total of 1,148 listed in the sampling frame. For the optimum three-stage design, between 500 and 600 PSUs would be selected. This would allow for greater stratification. In addition, the NAEP all-state requirement would not be applicable to the NLS design. Thus, the PSU stratification designed for NLS could be more effective than the Year 05 NAEP scheme. More effective stratification should improve the efficiency of the "best" three-stage design relative to the "best" two-stage design. It is doubtful however that this improvement would be sufficient to change the results appreciably. In conclusion, RTI does not feel that the gains in precision projected here for a three-stage design relative to the present two-stage design are sufficient to justify redesigning the sample.

References

[1] R. P. Moore, B. V. Shah, and R. E. Folsom, "Efficiency Study of NLS Base-Year Design Report on RTI Project 22U-884-3, Research Triangle Institute, Research Triangle Park, North Carolina, November, 1974.

[2] R. P. Moore and B. V. Shah, "Effects of Stratification, Clustering, and Unequal Weighting on the Variances of NLS Statistics", Report on Research Triangle Institute Project 22U-884-3, Research Triangle Institute, Research Triangle Park, North Carolina, February, 1975.

[3] R. P. Moore and B. V. Shah, "Choosing the "Optimum" Design for an NLS Cohort", Report on Research Triangle Institute Project 22U-884-21, Research Triangle Institute, Research Triangle Park, North Carolina.

[4] J. R. Chromy, A. F. Clemmer, R. E. Folsom, B. L. Jones, and J. E. Richardson, "Final Report on National Assessment of Educational Progress Sampling and Weighting Activities for Assessment Year 05", Prepared for National Assessment of Educational Progress, Research Triangle Institute, Research Triangle Park, North Carolina, October 1975.

[5] "Sample Design for the Selection of a Sample of Schools with Twelfth-Graders for a Longitudinal Study", unpublished manuscript, Westat, Inc., Rockville, Maryland, June 1972.

[6] B. V. Shah and C. N. Dillard, VARCOMP Program for Computing Variance Components, Working Paper No. 30 on RTI project 255U-796, Research Triangle Institute, Research Triangle Park, North Carolina, 1973.

APPENDIX A

SUMMARY STATEMENTS OF 1972 NLS AND 1974 NAEP DESIGN

APPENDIX A

1. The NLS Sample Design for the 1972 Cohort

The NLS 1972 sample design may be described as a deeply stratified two-stage probability sample with schools as first-stage sampling units and students (also guidance counselors) as second-stage units. The target population consisted of all 1972 twelfth graders enrolled in all public, private, and church-affiliated high schools in the 50 States and the District of Columbia. The first-stage sampling frame was constructed from computerized school files maintained by the United States Office of Education and by the National Catholic Educational Association.

The school sampling frame was stratified into 289 major strata based on the following variables:

- Type of control (public or nonpublic);
- Geographic region (Northeast, North Central, South, and West);
- Grade 12 enrollment (less than 300; 300 to 599; 600 or more);
- Proximity to institutions of higher learning (three categories);
- Percentage minority group enrollment (eight levels), and
- Income level of the community (11 categories for public schools, 8 categories for Catholic schools, and a single category for other schools).

Then, on the basis of the degree of urbanization, one or more final strata were defined within each major stratum to generate a total of 600 final strata.

In order to increase the numbers of disadvantaged students in the sample, schools located in low income areas and schools with high proportions of minority group enrollments were sampled at approximately twice the sampling rate used for the remaining schools. Schools in the smallest enrollment strata (less than 300 seniors) were selected with probabilities

proportional to their estimated numbers of senior students and without replacement. Schools in the remaining strata were selected with equal probabilities and without replacement. Within each final stratum, four schools were selected initially, and then two of the four were randomly designated as the primary selections. The other two schools were retained as backup or substitute selections and were used in the sample only if one or both of the primary schools did not cooperate (for example, closed, refused, or ineligible where ineligible relates here to schools for handicapped or legally confined students or that did not enroll students of their own). Five strata comprised exceptions to this methodology because they contained only three schools each. Samples of 18 students and two counselors per school were selected; five additional students and one additional counselor were selected as alternates. Students and counselors were selected from each sample school with equal probabilities and without replacement. Thus, the primary sample consisted initially of 1,200 schools (two per stratum), as many as 2,400 counselors (two per sample school), and as many as 21,600 students (18 per sample school).

2. The NAEP Year 05 (1974) Sample Design for In-school 17-year-old Students

The NAEP in-school sample was a stratified three-stage selection of PSUs (primary sampling units), schools within PSUs, and eligible 17-year-old students within schools. Stratification was carried out at the first two-stages in the sampling process.

The PSU sampling frame consisted of 243 SMSA (Standard Metropolitan Statistical Areas) and 905 non-SMSA county groups. Contiguous non-SMSA counties were combined into PSUs to meet minimum size requirements. The size measures used to select PSUs and to allocate the primary sample were

adjusted to overrepresent, at a rate of two-to-one, the low metropolitan and extreme rural areas.

One hundred and sixteen primary units were selected for the 1974 assessment. The sample PSUs were allocated to the four geographic regions (Northeast, Central, Southeast, and West) in proportion to their adjusted size measure. The geographic region designations are those used by the Office of Business Economics Department of Commerce. The 15 largest SMSAs had adjusted size measures large enough to warrant their inclusion in the sample with certainty. Two additional SMSA PSUs, namely Denver and Phoenix, became self-representing by virtue of the stratification scheme used to meet an all-state requirement imposed on the sample. Non-self-representing PSUs were selected with probabilities strictly proportional to their adjusted size measures. Besides stratifying by geographic region, a size of community variable was defined. Five basic size of community (SOC) strata were defined by:

SOC 1: The 17 self-representing PSUs;

SOC 2: The remaining 48 SMSAs with total populations in excess of 500,000;

SOC 3: The remaining 178 SMSAs;

SOC 4: Non-SMSA counties with 60 percent or less of their 14-year-old population classified as rural in the 1970 census;*

SOC 5: Non-SMSA counties with more than 60 percent of their 14-year-old population classified as rural in the 1970 census.

SOC was used as a guide to stratification rather than a major dimension of stratification. A detailed description of the primary stratification scheme used to meet the "all-state" representation requirement and simultaneously

---

*1970 14-year-olds were chosen to approximate 1974 17's instead of 1970 13-year-olds, since the single age totals for 13's were not available at the census tract level.

preserve as much control over size of community (SOC) as possible is presented in RTI's NAEP Year 05 Final Report [4].

Oversampling of the low SES metropolitan subpopulation was restricted to the largest 65 SMSAs (SOC 1 and SOC 2) whose total populations each exceeded 500,000. The extreme rural population was also oversampled. An extreme rural subpopulation was associated with the rural portion of counties whose 1970 14-year-old population was at least 75 percent rural. These extreme rural areas accounted for 10 percent of the 1970 14-year-old population. The low SES metropolitan areas contained 7.3 percent of the 14-year-olds.

Further efforts were made to control the sample distribution by size and type of community at the school selection and test package assignment phases.** The second stage involved selecting a stratified probability sample of schools within each of the 116 sample PSUs. An estimate of the number of age class eligibles was obtained for each school in the sample PSUs. A low SES metropolitan (SOC 1 and SOC 2) and an extreme rural (SOC 3) substratum was formed within the PSUs of the indicated SOC categories. Low SES metropolitan and extreme rural schools were oversampled at a rate of two-to-one in relation to nonextreme schools. Schools were selected with probabilities strictly proportional to the estimated number of age class eligibles within each extreme or nonextreme substratum. A total of 1,198 schools were selected for the 17-year-old assessment. A simple random sample of 16 eligible 17-year-old students was selected from each sample school per package assigned the sample school. Four of the 16

**National Assessment test exercises are packaged into subsets requiring no more than 50 minutes to complete. Generally, a student is assigned to only one such package. Group packages were assigned to 12 students and were administered simultaneously to the 12.

students were randomly designated as alternates. The total 17-year-old sample size, across all 16 group packages, was 36,709.

APPENDIX B

TABLES FOR DETERMINING COST COMPONENTS

Table B-1 Estimated Direct Charges Allocated to School and Student Costs, NLS Base-Year Survey, Two-Stage Design

Item	School cost (C_1) dollars	Student cost (C_2) dollars
Printing	5,280	21,120
Postage and shipping	3,700	14,800
Telephone and communications	14,000	-
School honorariums	30,000	30,000
Data processing:		
Selecting student samples	1,300	-
Editing and keytaping	3,570	32,130
Total Direct Charges	57,850	98,050

Table B-2 Estimated Man-months of Direct Labor Allocated to School and Student Costs, NLS Base-Year Survey, Two-Stage Design

Item	School cost (C_1)		Student cost (C_2)	
	Professional man-months	Support man-months	Professional man-months	Support man-months
Task 2	2.0	7.0	-	-
Task 3	5.0	10.0	-	-
Task 4	-	-	-	-
Task 5	7.0	9.0	-	-
Task 6	.7	3.5	6.3	31.5
Task 7	.8	4.5	7.2	40.5
Total	15.5	34.0	13.5	72.0

Table B-3 Calculation of Estimated Variable Cost Parameters C_1 and C_2, NLS

Base-Year Survey, Two-Stage Design

Item	School cost (C_1) dollars	Student cost (C_2) dollars
Direct labor		
Professional	23,250	20,250
Support	20,400	43,200
Total	43,650	63,450
Indirect technical expense[1/]	32,738	47,588
Direct charges	57,850	98,050
Overall management and service expense[2/]	14,766	23,000
Total estimated costs	149,004	232,088
Fixed fee[3/]	7,450	11,604
Total estimated cost plus fixed fee	156,454	243,692
Estimated cost parameters[4/]	130.38	11.28

[1/] Computed as 75 percent of direct labor.

[2/] Computed as 11 percent of direct labor, indirect technical expense, and direct charges.

[3/] Calculated as five percent of total estimated costs.

[4/] Computed by dividing the total estimated school cost plus fee by 1,200 and corresponding student cost plus fee by 21,600.

Table B-4
Estimated Direct Charges Allocated To
PSU, School, and Student Costs

A. Three-Stage Design

Item	PSU-Cost $\$_1$ dollars	Student-Cost $\$_3$ dollars
Printing subcontract		86,940
Other reproductions		1,750
Materials		4,150
Communications and shipping		88,730
Services		4,225
Temporary labor		104,089
Staff travel	5,549*	22,477*
Interviewer salaries	12,468*	50,504*
Interviewer mileage	9,865*	39,959*
Total direct charges	27,882	402,824

The school-cost is assumed to be zero.

*Three-stage model Interviewer's salary and mileage is 69.2% of the two-stage model. These expenses and staff travel are allocated 19.8% for PSU-cost and 80.2% for student-cost.

B. Two-Stage Design

Item	School-Cost C_1 dollars	Student-Cost C_2 dollars
Printing subcontract		86,940
Other reproductions		1,750
Materials		4,150
Communications and shipping		88,730
Services		4,225
Temporary labor		104,089
Staff travel	12,150*	28,350*
Interviewer salaries	27,300*	63,700*
Interviewer mileage	21,600*	50,400*
Total direct charges	61,050	432,334

*These expenses and staff travel are allocated as 30% school-costs and 70% student-costs.

Table B-5
Calculation of Estimated Variable Cost Parameters, C_1 and C_2 for an NLS Follow-Up Survey Using a Two-Stage Sampling Design
Cost Procedure B

Item	School-Cost C_1 dollars	Student-Cost C_2 dollars
Direct labor		
Survey specialists	16,524	38,556
Research assistants		34,310
Survey clerks		14,438
Total	16,524	87,304
Indirect technical expense[1]	12,393	65,478
Direct charges	61,050	432,334
Overall management and service expense[2]	9,896	64,363
Total estimated costs	99,863	649,479
Fixed fee[3]	4,993	32,474
Total estimated cost plus fixed fee	104,856	681,953
Number of schools and students assumed selected	1,200	21,600
Estimated cost parameters[4]	87.38	31.57

[1] Computed as 75 percent of direct labor.

[2] Computed as 11 percent of direct labor, indirect technical expense, and direct charges.

[3] Calculated as five percent of total estimated costs.

[4] Calculated as total estimated cost plus fixed fee divided by the number selected.

Table B-6
Calculation of Estimated Variable Cost Parameters, $\$_1$, $\$_2$, and $\$_3$ for an NLS Follow-Up Survey Using a Three-Stage Sampling Design,*
Cost Procedure B

Item	PSU-Cost $\$_1$ dollars	Student-Cost $\$_2$ dollars
Direct labor		
Survey Specialists	7,547	30,568
Research Assistants		34,310
Survey clerks		14,438
Total	7,547	79,316
Indirect technical expense[1/]	5,660	59,487
Direct charges	27,882	402,824
Overall management and service expense[2/]	4,520	59,579
Total estimated costs	45,609	601,206
Fixed fee[3/]	2,280	30,060
Total estimated cost plus fixed fee	47,889	631,266
Number of PSUs and students assumed selected	400	21,600
Estimated cost parameters[4/]	119.72	29.23

1/ Computed as 75 percent of direct labor.

2/ Computed as 11 percent of direct labor, indirect technical expense, and direct charges.

3/ Calculated as five percent of total estimated costs.

4/ Total estimated cost plus fixed fee divided by the number selected.

* The school cost is assumed to be zero.

Table B-7 Estimated Direct Charges Allocated to School and Student Costs, NLS Follow-Up Survey, Two-Stage Design

Item	School cost (C_1) dollars	Student cost (C_2) dollars
Printing subcontract	-	86,940
Other reproductions	-	1,750
Materials	-	4,150
Communications and shipping	-	88,730
Services	-	4,225
Temporary labor	-	104,089
Staff travel	-	40,500
Interviewer salaries	27,300	63,700
Interviewer mileage	21,600	50,400
Total direct charges	48,900	444,484

Table B-8 Estimated Man-Months of Direct Labor, by Task, by Position, NLS Follow-Up Survey, Two-Stage Design

Item	Survey specialists man-months	Research assistants man-months	Survey clerks man-months
Task 2	1.0	8.00	1.00
Task 3	5.0	30.00	16.25
Task 7	45.0	9.0	9.0
Total	51.0	47.0	26.25

Table B-9 Calculation of Estimated Variable Cost Parameters, C_1 and C_2, for an NLS Follow-Up Survey, Two-Stage Design

Item	School cost (C_1) dollars	Student cost (C_2) dollars
Direct labor		
Survey specialists	-	55,080
Research assistants	-	34,310
Survey clerks	-	14,438
Total		103,828
Indirect technical expense[1/]	-	77,871
Direct charges	48,900	444,484
Overall management and service expense[2/]	5,379	68,880
Total estimated costs	54,279	695,063
Fixed fee[3/]	2,714	34,753
Total estimated cost plus fixed fee	56,993	729,816
Estimated cost parameters	47.49	33.79

1/ Computed as 75 percent of direct labor

2/ Computed as 11 percent of direct labor, indirect technical expense, and direct charges

3/ Calculated as five percent of total estimated costs.

4/ Calculated by dividing total school cost plus fee by 1,200 and total student cost plus fee by 21,600.